(19)日本国特許庁（ＪＰ）　(12) **公開特許公報（Ａ）**　(11)特許出願公開番号
特開2000－305832
（P2000－305832A）
(43)公開日　平成12年11月２日(2000.11.2)

(51)Int.Cl.'	識別記号	ＦＩ	テーマコード'(参考)
Ｇ０６Ｆ　12/00	５３５	Ｇ０６Ｆ　12/00	５３５Ｚ　５Ｂ０４５
	５３１		５３１Ｒ　５Ｂ０８２
9/46	３６０	9/46	３６０Ｄ　５Ｂ０９８
15/16	６４０	15/16	６４０Ａ

審査請求　有　　請求項の数９　ＯＬ　（全９頁）

(21)出願番号　特願平11－116853

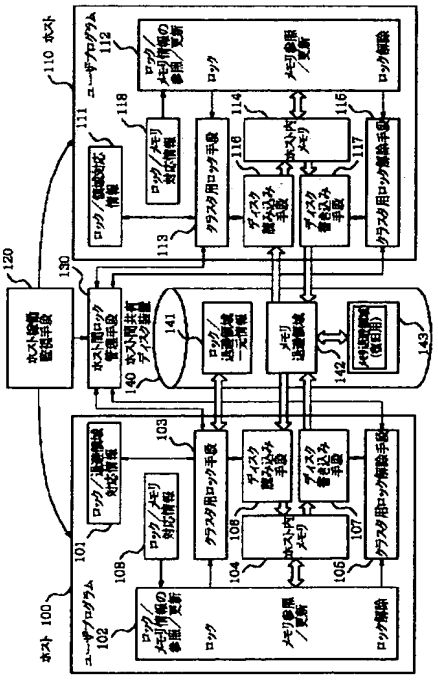
(22)出願日　平成11年４月23日(1999.4.23)

(71)出願人　000164449
九州日本電気ソフトウェア株式会社
福岡市早良区百道浜２丁目４－１　ＮＥＣ九州システムセンター
(72)発明者　大塚　英明
福岡県福岡市早良区百道浜２－４－１　九州日本電気ソフトウェア株式会社内
(74)代理人　100082935
弁理士　京本　直樹　（外２名）

Ｆターム(参考)　5B045 EE03 EE19 GG01
5B082 DA02 DC06 DE06 FA17
5B098 AA10 DD01 GB05 GD16 GD21
GD22

(54)【発明の名称】　複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法

(57)【要約】
【課題】複数ホストから構成されるクラスタシステムにおけるメモリの共有を行う場合、ユーザプログラムの改装が必要。
【解決手段】ホスト１００はユーザプログラム１０２とクラスタ用ロック手段１０３とホスト内メモリ１０４とクラスタ用ロック解除手段１０５とディスク読み込み手段１０６とディスク書き込み手段１０７とロック／メモリ対応情報１０８とロック／退避領域対応情報１０１を有しホスト１１０も同機能があり複数ホスト間の共用機能としてホスト稼動監視手段１２０とホスト間ロック管理手段１３０とホスト間共有ディスク装置１４０がありクラスタ用ロック手段とクラスタ用ロック解除手段実行時に自動的にホスト内メモリを更新することにより１ホスト内で動作していた複数プロセスからメモリを共有するユーザプログラムを改装なしで複数ホストから構成されるクラスタシステムで実行可能にする。



1

【特許請求の範囲】

【請求項１】　複数プロセスからメモリを共有するユーザプログラムと、クラスタ用ロック手段と、ホスト内メモリと、クラスタ用ロック解除手段と、ディスク読み込み手段と、ディスク書き込み手段と、ロック／メモリ対応情報と、ロック／退避領域対応情報とを備えたホストと、複数ホスト間にホスト稼動監視手段と、ホスト間ロック管理手段と、ホスト間共有ディスク装置とを備え、クラスタ用ロック手段およびクラスタ用ロック解除手段実行時、自動的にホスト内メモリを更新することにより、ホスト内で動作していた複数プロセスからメモリを共有するユーザプログラムを、ユーザプログラムの改装なしで複数ホストから構成されるクラスタシステムで実行可能にすることを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有装置。

【請求項２】　ホスト稼動監視手段により各ホストの稼動状況を監視し、ホスト間ロック管理手段によりホスト間でロック制御を行い、ホスト間共有ディスク装置はロック識別子とメモリ退避領域を管理するロック／退避領域一元情報とロック識別子毎に存在するメモリ退避領域およびメモリ退避領域（復旧用）から構成し、各ホスト毎に存在するロック／メモリ対応情報はロック識別子とそのロック識別子に対応する共有メモリであるホスト内メモリのメモリアドレスとホスト内メモリのメモリサイズを持ちロック識別子をキーに情報が管理され、各ホスト毎に存在するロック／退避領域対応情報およびホスト間共有ディスク装置上に記録されるロック／退避領域一元情報はロック識別子とそのロック識別子に対応する共有メモリの退避領域であるメモリ退避領域の情報を管理する退避領域情報とメモリ退避領域（復旧用）の情報を管理する退避領域情報（復旧用）を持ちロック識別子をキーに情報が管理され、ホスト間共有ディスク装置のメモリ退避領域はホスト間共有ディスク装置へのレコード書き込み中にホストダウンによりアクセス障害が発生した場合のレコードの正当性のチッェクをレコードの先頭部の書き込みチェック領域Ａと終端部の書き込みチェック領域Ｂのデータが正しければその間のメモリ退避領域の値も正しい事が保証されるレコードフォーマットにより行うよう構成されることを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有装置。

【請求項３】　ホスト間共有ディスク装置へのレコード書き込み中にホストダウンによりアクセス障害が発生した場合のレコードの正当性のチッェクをレコードの先頭部と終端部のデータが正しければその間の値も正しい事が保証されるレコードフォーマットにより行うことを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有装置。

【請求項４】　ロック／メモリ対応情報は、ホスト内で最初にユーザプログラムが実行された時に、ホスト内メモリの領域確保やロック識別子の取得と同期して作成され、ロック／退避領域対応情報は、クラスタ用ロック手段実行時に参照され、ユーザプログラムからロックが要求されたロック識別子に対応する情報が登録されていなければ、ロック／退避領域一元情報を参照し、登録されていなければ、新たにメモリ退避領域およびメモリ退避領域（復旧用）をそのロック識別子用に確保し、その領域をアクセスするための情報をロック／退避領域一元情報とロック／退避領域対応情報に登録し、メモリ退避領域およびメモリ退避領域（復旧用）は領域の確保時に初期化し、クラスタ用ロック手段は、ホスト間ロック管理手段に対して、ロックを要求しロック成功後に、ロック／退避領域対応情報とロック／メモリ対応情報を参照し、ディスク読み込み手段を利用して、メモリ退避領域からホスト内メモリにデータを読み込み、ユーザプログラムはロック後に、ホスト内メモリの参照および更新を行い、クラスタ用ロック解除手段は、ロック／退避領域対応情報とロック／メモリ対応情報を参照し、ディスク書き込み手段を利用して、メモリ退避領域にホスト内メモリのデータを書き込んだ後、ホスト間ロック管理手段に対して、ロック解除を要求し、ホスト間ロック管理手段は、各ホストからのロック要求を管理するとともに、あるホストがダウンした場合、ホスト稼動監視手段から通知を受け、ダウンしたホストが取得していたロックを解除することを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有方法。

【請求項５】　ユーザプログラムがホスト内で最初に実行された時、ホスト内メモリの領域確保やロック識別子の取得を行い、ロック／メモリ対応情報に登録し、ユーザプログラムが実行されホスト内メモリを参照または更新時にロックを行うために、クラスタ用ロック手段が実行され、クラスタ用ロック手段は、ホスト間ロック管理手段と連携し、クラスタシステム全体のロックが成功した場合、ユーザプログラムが要求したロックのモードにより、参照ロックの場合、参照ロックは、共有メモリを排他的に参照するため、ユーザプログラムからのロック要求時にクラスタ用ロック手段はロック／メモリ対応情報とロック／退避領域対応情報を参照し、ロック識別子に対応するメモリアドレスおよびメモリサイズ、退避領域情報、退避領域情報（復旧用）を取得し、ディスク読み込み手段を使用してホスト間共有ディスク装置のメモリ退避領域から、ホスト内メモリにデータを読み込み、メモリ退避領域のレコードの先頭部の書き込みチェック領域とレコードの終端部の書き込みチェック領域をチェックし、ホストダウンによりデータが不正な場合は、メモリ退避領域（復旧用）から、ホスト内メモリにデータを読み込み、メモリ退避領域（復旧用）の情報をメモリ退避領域に書き込んだ後、ユーザプログラムに制御を戻し、ユーザプログラムから参照ロックのロック解除要求時に、クラスタ用ロック解除手段はホスト間ロック管理

手段と連携し、ロック解除を行い、更新ロックの場合、更新ロックは、共有メモリを排他的に更新するため、ユーザプログラムからのロック要求時にクラスタ用ロック手段はロック／メモリ対応情報とロック／退避領域対応情報を参照し、ロック識別子に対応するメモリアドレスおよびメモリサイズ、退避領域情報、退避領域情報（復旧用）を取得し、ディスク読み込み手段を使用してホスト間共有ディスク装置のメモリ退避領域から、ホスト内メモリにデータを読み込み、メモリ退避領域のレコードの先頭部の書き込みチェック領域とレコードの終端部の書き込みチェック領域をチェックし、ホストダウンによりデータが不正な場合は、メモリ退避領域（復旧用）から、ホスト内メモリにデータを読み込み、メモリ退避領域（復旧用）の情報をメモリ退避領域に書き込んだ後、ユーザプログラムに制御を戻、メモリ退避領域のデータが正しい場合はメモリ退避領域の内容をディスク書き込み手段を使用してメモリ退避領域（復旧用）に書き込み、ユーザプログラムから更新ロックのロック解除要求時に、クラスタ用ロック解除手段はディスク書き込み手段を使用してホスト内メモリをホスト間共有ディスク装置のメモリ退避領域に書き込み、ホスト間ロック管理手段と連携し、ロック解除を行うことを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有方法。

【請求項６】　あるホストでロック中にホストダウンが発生した場合、ホスト内メモリの領域確保やロック識別子の取得を行い、ロック／メモリ対応情報に登録し、ユーザプログラムが実行されホスト内メモリを参照または更新時にロックを行うために、クラスタ用ロック手段が実行され、クラスタ用ロック手段は、ホスト間ロック管理手段と連携し、クラスタシステム全体のロックが成功した場合、ユーザプログラムが要求したロックのモードにより、参照ロックの場合、参照ロックは、共有メモリを排他的に参照するため、メモリに対する復旧処理は不要で、ロックに関しては、ホストでホストダウンが発生した場合、ホスト稼動監視手段は、ホストのダウンを認識すると、ホスト間ロック管理手段に通知し、ホスト間ロック管理手段はホストからのロックをすべて解除し、ロック解除により動作可能となったユーザプログラムからのロック要求時にクラスタ用ロック手段はロック／メモリ対応情報とロック／退避領域対応情報を参照し、ロック識別子に対応するメモリアドレスおよびメモリサイズ、退避領域情報、退避領域情報（復旧用）を取得し、ディスク読み込み手段を使用してホスト間共有ディスク装置のメモリ退避領域から、ホスト内メモリにデータを読み込み、メモリ退避領域のレコードの先頭部の書き込みチェック領域とレコードの終端部の書き込みチェック領域をチェックし、ホストダウンによりデータが不正な場合は、メモリ退避領域（復旧用）から、ホスト内メモリにデータを読み込み、メモリ退避領域（復旧用）の情報をメモリ退避領域に書き込んだ後、ユーザプログラムに制御を戻し、ユーザプログラムから参照ロックのロック解除要求時に、クラスタ用ロック解除手段はホスト間ロック管理手段と連携し、ロック解除を行い、更新ロックの場合更新ロックは、共有メモリを排他的に更新するため、メモリに対する復旧処理は不要で、ロック解除により動作可能となったユーザプログラムからのロック要求時にクラスタ用ロック手段はロック／メモリ対応情報とロック／退避領域対応情報を参照し、ロック識別子に対応するメモリアドレスおよびメモリサイズ、退避領域情報、退避領域情報（復旧用）を取得し、ディスク読み込み手段を使用してホスト間共有ディスク装置のメモリ退避領域から、ホスト内メモリにデータを読み込み、メモリ退避領域のレコードの先頭部の書き込みチェック領域とレコードの終端部の書き込みチェック領域をチェックし、ホストダウンによりデータが不正な場合は、メモリ退避領域（復旧用）から、ホスト内メモリにデータを読み込み、メモリ退避領域（復旧用）の情報をメモリ退避領域に書き込んだ後、ユーザプログラムに制御を戻、メモリ退避領域のデータが正しい場合はメモリ退避領域の内容をディスク書き込み手段を使用してメモリ退避領域（復旧用）に書き込み、ユーザプログラムから更新ロックのロック解除要求時に、クラスタ用ロック解除手段はディスク書き込み手段を使用してホスト内メモリをホスト間共有ディスク装置のメモリ退避領域に書き込み、ホスト間ロック管理手段と連携し、ロック解除を行うことを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有方法。

【請求項７】　各ホストのユーザプログラムが全て参照ロックの場合、ホストのユーザプログラムは、ロック／対応情報を参照し、共有メモリを参照するためのロック（参照ロック）を行い、ホスト間共有ディスク装置中のメモリ退避領域からホスト内メモリにデータが読み込み、他のホストのユーザプログラムは自己のロック／対応情報を参照し、共有メモリを参照するためのロック（参照ロック）を行い、参照ロックであるためロックに成功し、ホスト間共有ディスク装置中のメモリ退避領域から自己のホスト内メモリにデータが読み込むことを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有方法。

【請求項８】　参照ロックのユーザプログラムと更新ロックのユーザプログラムの場合、参照ロックのホストのユーザプログラムは、ロック／対応情報を参照し、共有メモリを参照するためのロック（参照ロック）を行い、ホスト間共有ディスク装置中のメモリ退避領域からホスト内メモリにデータが読み込み、更新ロックのホストのユーザプログラムは自己のロック／対応情報を参照し、共有メモリを更新するためのロック（更新ロック）を行うことによりユーザプログラムで参照ロック中であるためロック待ち状態となり、ユーザプログラムでロックが

解除された後でロックが成功し、ホスト間共有ディスク装置中のメモリ退避領域からホスト内メモリにデータが読み込まれ、更新ロックのユーザプログラムでロック解除時に、自己のホスト内メモリは、ホスト間共有ディスク装置中のメモリ退避領域に書き込まれることを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有方法。

【請求項９】　各ホストのユーザプログラムが全て更新ロックの場合、ホストのユーザプログラムは、ロック／対応情報を参照し、共有メモリを更新するためのロック（更新ロック）を行い、ホスト間共有ディスク装置中のメモリ退避領域からホスト内メモリにデータが読み込み、他のホストのユーザプログラムは自己のロック／対応情報を参照し、共有メモリを更新するためのロック（更新ロック）を行うことによりユーザプログラムで更新ロック中であるためロック待ち状態となり、ユーザプログラムがロック解除時に、ホスト内メモリは、ホスト間共有ディスク装置中のメモリ退避領域に書き込まれ、ロックが解除された後で他のユーザプログラムでロックが成功し、ホスト間共有ディスク装置中のメモリ退避領域からホスト内メモリにデータが読み込まれ、他のユーザプログラムでロック解除時に、自己のホスト内メモリは、ホスト間共有ディスク装置中のメモリ退避領域に書き込まれることを特徴とする複数ホストから構成されるクラスタシステムにおけるメモリの共有方法。

【発明の詳細な説明】

【０００１】

【発明の属する技術分野】本発明は複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法に関し、特に１ホスト内で動作実績がある複数プロセスからメモリを共有するユーザプログラムを、ユーザプログラムの改装なしで複数ホストから構成されるクラスタシステムで実行可能とする複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法に関する。

【０００２】

【従来の技術】従来のメモリ共有型のマルチプロセスシステムの場合のメモリ共有方式は、特開平０８－１６６９３３号公報に記載されている。この従来のメモリ共有方式は、各ホストに存在するホスト間共有データ領域と、ホスト間共有データ領域所在管理手段と、データ送受信手段から構成されている。

【０００３】このような構成を有する従来のメモリ共有方式はつぎのように動作する。

【０００４】すなわち、プログラムがホスト間共有データ領域をアクセスすると、ホスト間共有データ領域管理手段が作動し、他ホストと通信を行い、ホスト間共有データ領域を補正している。

【０００５】

【発明が解決しようとする課題】上述した従来のメモリ共有方式は、第１の問題点は、すでに１ホスト上で動作していた既存ユーザプログラムの改造が必要であるということである。

【０００６】その理由は、ホスト間共有データ領域をアクセスするように改造しなければならないためである。

【０００７】第２の問題点は、処理性能が遅いということである。

【０００８】その理由は、他ホストとの通信が必要であり、ホスト台数に比例して通信量が増加するためである。また、メモリを排他的にアクセスしたい場合でも、他ホストへの確認が必要になる点である。

【０００９】

【課題を解決するための手段】本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法は、複数プロセスからメモリを共有するユーザプログラムと、クラスタ用ロック手段と、ホスト内メモリと、クラスタ用ロック解除手段と、ディスク読み込み手段と、ディスク書き込み手段と、ロック／メモリ対応情報と、ロック／退避領域対応情報とを備えたホストと、複数ホスト間にホスト稼動監視手段と、ホスト間ロック管理手段と、ホスト間共有ディスク装置とを備え、ホスト稼動監視手段により各ホストの稼動状況を監視し、ホスト間ロック管理手段によりホスト間でロック制御を行い、ホスト間共有ディスク装置はロック識別子とメモリ退避領域を管理するロック／退避領域一元情報とロック識別子毎に存在するメモリ退避領域およびメモリ退避領域（復旧用）から構成し、各ホスト毎に存在するロック／メモリ対応情報はロック識別子とそのロック識別子に対応する共有メモリであるホスト内メモリのメモリアドレスとホスト内メモリのメモリサイズを持ちロック識別子をキーに情報が管理され、各ホスト毎に存在するロック／退避領域対応情報およびホスト間共有ディスク装置上に記録されるロック／退避領域一元情報はロック識別子とそのロック識別子に対応する共有メモリの退避領域であるメモリ退避領域の情報を管理する退避領域情報とメモリ退避領域（復旧用）の情報を管理する退避領域情報（復旧用）を持ちロック識別子をキーに情報が管理され、ホスト間共有ディスク装置のメモリ退避領域はホスト間共有ディスク装置へのレコード書き込み中にホストダウンによりアクセス障害が発生した場合のレコードの正当性のチェックをレコードの先頭部の書き込みチェック領域Ａと終端部の書き込みチェック領域Ｂのデータが正しければその間のメモリ退避領域の値も正しい事が保証されるレコードフォーマットにより行うよう構成され、ロック／メモリ対応情報は、ホスト内で最初にユーザプログラムが実行された時に、ホスト内メモリの領域確保やロック識別子の取得と同期して作成され、ロック／退避領域対応情報は、クラスタ用ロック手段実行時に参照され、ユーザプログラムからロックが要求されたロック識別子に対応する情報が登録されていなければ、ロ

ック／退避領域一元情報を参照し、登録されていなければ、新たにメモリ退避領域およびメモリ退避領域（復旧用）をそのロック識別子用に確保し、その領域をアクセスするための情報をロック／退避領域一元情報とロック／退避領域対応情報に登録し、メモリ退避領域およびメモリ退避領域（復旧用）は領域の確保時に初期化し、クラスタ用ロック手段は、ホスト間ロック管理手段に対して、ロックを要求しロック成功後に、ロック／退避領域対応情報とロック／メモリ対応情報を参照し、ディスク読み込み手段を利用して、メモリ退避領域からホスト内メモリにデータを読み込み、ユーザプログラムはロック後に、ホスト内メモリの参照および更新を行い、クラスタ用ロック解除手段は、ロック／退避領域対応情報とロック／メモリ対応情報を参照し、ディスク書き込み手段を利用して、メモリ退避領域にホスト内メモリのデータを書き込んだ後、ホスト間ロック管理手段に対して、ロック解除を要求し、ホスト間ロック管理手段は、各ホストからのロック要求を管理するとともに、あるホストがダウンした場合、ホスト稼動監視手段から通知を受け、ダウンしたホストが取得していたロックを解除し、クラスタ用ロック手段およびクラスタ用ロック解除手段実行時、自動的にホスト内メモリを更新することにより、ホスト内で動作していた複数プロセスからメモリを共有するユーザプログラムを、ユーザプログラムの改装なしで複数ホストから構成されるクラスタシステムで実行可能にするように構成されている。

【００１０】

【発明の実施の形態】本発明は、１ホスト内で動作実績がある複数プロセスからメモリを共有するユーザプログラムを、ユーザプログラムの改装なしで複数ホストから構成されるクラスタシステムで実行可能とするための方式である。ここで言うクラスタシステムとは、同一構成のシステムを複数台のホストに実装し、各ホストの負荷状況に応じて、処理を自動的に負荷分散するシステムのことである。

【００１１】次に、本発明の実施の形態について図面を参照して説明する。

【００１２】図１は本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法の一実施の形態を示すブロック図である。

【００１３】図１を参照すると、本発明の実施の形態はホスト１００と、ホスト１１０と、各ホストの稼動を監視するホスト稼動監視手段１２０と、ホスト間のロック制御を行うホスト間ロック管理手段１３０と、ホスト間で共有するホスト間共有ディスク装置１４０から構成されている。

【００１４】クラスタシステムを構成するホスト１００には、ユーザプログラム１０２と、ロック機能のユーザインタフェースを提供するクラスタ用ロック手段１０３と、ユーザプログラム１０２が他プログラムと共有するメモリ領域であるホスト内メモリ１０４と、ロック解除機能のユーザインタフェースを提供するクラスタ用ロック解除手段１０５と、ホスト間共有ディスク装置１４０からデータをホスト内メモリ１０４に読み込むディスク読み込み手段１０６と、ホスト内メモリ１０４からホスト間共有ディスク装置１４０にデータを書き込むディスク書き込み手段１０７と、ロック識別子と共有メモリ領域の対応を管理するロック／メモリ対応情報１０８と、ロック識別子とメモリの退避領域の対応を管理するロック／退避領域対応情報１０１から構成される。

【００１５】ホスト１１０は、ホスト１００と同じ構成であり、図１ではホスト２台の構成となっているが、ホスト台数は自由に増やすことができる。

【００１６】ホスト稼動監視手段１２０は、各ホストの稼動状況を監視するために使用される。

【００１７】ホスト間ロック管理手段１３０は、ホスト間でロック制御を行うために使用される。

【００１８】ホスト間共有ディスク装置１４０は、ロック識別子やメモリ退避領域を管理するロック／退避領域一元情報１４１と、ロック識別子毎に存在するメモリ退避領域１４２およびメモリ退避領域（復旧用）１４３から構成される。

【００１９】各ホスト毎に存在するロック／メモリ対応情報１０８には、図３の本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるロック識別子と、そのロック識別子に対応する共有メモリであるホスト内メモリのメモリアドレスと、メモリサイズとの対応情報の構造を示す図に示すように、ロック識別子３０１と、そのロック識別子に対応する共有メモリであるホスト内メモリ１０４のメモリアドレス３０２と、ホスト内メモリ１０４のメモリサイズ３０３を持ち、ロック識別子３０１をキーに情報が管理される。

【００２０】各ホスト毎に存在するロック／退避領域対応情報１０１およびホスト間共有ディスク装置１４０上に記録されるロック／退避領域一元情報１４１には、図４の本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるロック識別子と、そのロック識別子に対応する共有メモリの退避領域情報と、退避領域情報（復旧用）との対応情報の構造を示す図に示すように、ロック識別子４０１と、そのロック識別子に対応する共有メモリの退避領域であるメモリ退避領域１４２の情報を管理する退避領域情報４０２と、メモリ退避領域（復旧用）１４３の情報を管理する退避領域情報（復旧用）４０３を持ち、ロック識別子４０１をキーに情報が管理される。

【００２１】ホスト間共有ディスク装置１４０のメモリ退避領域１４２は、図５の本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるホスト間共有ディスク装置のメモリ退避領域

の書き込みチェック領域と、メモリ退避領域と、書き込みチェック領域との対応情報の構造を示す図に示すように、書き込みチェック領域Ａ５０１と、メモリ退避領域５０２と、書き込みチェック領域Ｂ５０３から構成される。

【００２２】これらの手段はそれぞれつぎのように動作する。なお、説明はホスト１００およびホスト間で共有する手段等について行うが、ホスト１１０の各手段も同じ動作をする。

【００２３】ロック／メモリ対応情報１０８は、ホスト１００内で最初にユーザプログラム１０２が実行された時などに、ホスト内メモリ１０４の領域確保やロック識別子の取得と同期して作成される。

【００２４】ロック／退避領域対応情報１０１は、クラスタ用ロック手段１０３実行時に参照され、ユーザプログラム１０２からロックが要求されたロック識別子に対応する情報が登録されていなければ、ロック／退避領域一元情報１４１を参照し、そこにも登録されていなければ、新たにメモリ退避領域１４２およびメモリ退避領域（復旧用）１４３をそのロック識別子用に確保し、その領域をアクセスするための情報をロック／退避領域一元情報１４１とロック／退避領域対応情報１０１に登録する。

【００２５】メモリ退避領域１４２およびメモリ退避領域（復旧用）１４３は領域の確保時に初期化する。

【００２６】クラスタ用ロック手段１０３は、図２の本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるクラスタシステムを構成する１ホストであるメモリ共有型のマルチプロセスシステムを示すブロック図に示すメモリ共有型マルチプロセッサシステムであるホスト２００上の、ロック手段２０５と互換性があるユーザインタフェースを持ち、ホスト間ロック管理手段１３０に対して、ロックを要求しロック成功後に、ロック／退避領域対応情報１０１とロック／メモリ対応情報１０８を参照し、ディスク読み込み手段１０６を利用して、メモリ退避領域１４２からホスト内メモリ１０４にデータを読み込む。

【００２７】ユーザプログラム１０２はロック後に、ホスト内メモリ１０４の参照および更新を行う。

【００２８】ユーザプログラム１０２は、図２で示されるメモリ共有型のマルチプロセッサシステム上で動作するユーザプログラム２０２およびユーザプログラム２０３と同一のプログラムである。

【００２９】クラスタ用ロック解除手段１０５は、図２で示すメモリ共有型マルチプロセッサシステムであるホスト２００上の、ロック解除手段２０７と互換性があるユーザインタフェースを持ち、ロック／退避領域対応情報１０１とロック／メモリ対応情報１０８を参照し、ディスク書き込み手段１０７を利用して、メモリ退避領域１４２にホスト内メモリ１０４のデータを書き込んだ後

で、ホスト間ロック管理手段１３０に対して、ロック解除を要求する。

【００３０】ホスト間ロック管理手段１３０は、各ホストからのロック要求を管理し、表１のロック制御例に示すようなサービスを提供する。

【００３１】

【表１】

	参照ロック	更新ロック
参照ロック	同時実行可能	同時実行不可能
更新ロック	同時実行不可能	同時実行不可能

【００３２】また、ホスト間ロック管理手段１３０は、あるホストがダウンした場合は、ホスト稼動監視手段１２０から通知を受け、ダウンしたホストが取得していたロックを解除する。

【００３３】ホスト間共有ディスク装置１４０の特性として、レコード書き込み中にホストダウン等によりアクセス障害が発生した場合、レコードの先頭部と終端部のデータが正しければその間の値も正しい事が保証される場合のレコードフォーマットを図５に示している。実際に使用するディスク装置の特性にあったレコードフォーマットに変更する必要がある。

【００３４】メモリ退避領域（復旧用）１４３は、ホストダウン等によりメモリ退避領域１４２へのデータ書き込みに失敗した場合に、メモリ退避領域１４２を復旧するために使用される。

【００３５】次に、本発明の実施の形態の動作について、図１～図５および表１を参照して詳細に説明する。

【００３６】ます初めに、図２によりクラスタシステムを構成する１ホストであるメモリ共有型のマルチプロセスシステム上における従来動作を説明する。

【００３７】ホスト２００で動作するユーザプログラム２０２または２０３は、ホスト内で最初に実行された時などにホスト内メモリ２０６の領域確保やロック識別子の取得を行い、ロック／メモリ対応情報２０４に登録する。

【００３８】ロック／メモリ対応情報２０４は、ユーザプログラム２０２または２０３から参照され、ホスト内メモリ２０６アクセス時のロックやメモリアドレス等の取得に使用される。

【００３９】ユーザプログラム２０２または２０３からロック要求が行われると、ロック手段２０５は、ホスト内ロック管理手段２０１を使用してロックを行う。ホスト内ロック管理手段２０１は表１に示すようなサービスを提供する。

【００４０】ユーザプログラム２０２または２０３は、ロック後にホスト内メモリ２０６のアクセスを行う。

11

【００４１】ユーザプログラム２０２または２０３からロック解除要求が行われると、ロック解除手段２０７は、ホスト内ロック管理手段２０１を使用してロック解除を行う。

【００４２】次にクラスタシステムの場合の動作について説明する。

【００４３】ユーザプログラム１０２はホスト１００内で最初に実行された時などにホスト内メモリ１０８の領域確保やロック識別子の取得を行い、ロック／メモリ対応情報１０８に登録する。

【００４４】ユーザプログラム１０２が実行されホスト内メモリ１０４を参照または更新時にロックを行うために、クラスタ用ロック手段１０３が実行される。クラスタ用ロック手段１０３は、ホスト間ロック管理手段１３０と連携し、クラスタシステム全体のロックが成功した場合、ユーザプログラム１０２が要求したロックのモードにより、次の動作を行う。

【００４５】（１１）参照ロックの場合

参照ロックは、共有メモリを排他的に参照することが目的であるため、次のように動作する。

【００４６】ユーザプログラム１０２からのロック要求時にクラスタ用ロック手段１０３はロック／メモリ対応情報１０８とロック／退避領域対応情報１０１を参照し、図３および図４で示すロック識別子に対応するメモリアドレス３０２およびメモリサイズ３０３、退避領域情報４０２、退避領域情報（復旧用）４０３を取得し、ディスク読み込み手段１０６を使用してホスト間共有ディスク装置１４０のメモリ退避領域１４２から、ホスト内メモリ１０４にデータを読み込む。この時、メモリ退避領域１４２中の図５に示す書き込みチェック領域Ａ５０１と書き込みチェック領域Ｂ５０３をチェックし、ホストダウン等によりデータが不正な場合は、メモリ退避領域（復旧用）１４３から、ホスト内メモリ１０４にデータを読み込み、メモリ退避領域（復旧用）１４３の情報をメモリ退避領域１４２に書き込んだ後でユーザプログラム１０２に制御を戻す。

【００４７】ユーザプログラム１０２から参照ロックのロック解除要求時に、クラスタ用ロック解除手段１０５はホスト間ロック管理手段１３０と連携し、ロック解除を行う。

【００４８】（１２）更新ロックの場合

更新ロックは、共有メモリを排他的に更新することが目的であるため、次のように動作する。

【００４９】ユーザプログラム１０２からのロック要求時にクラスタ用ロック手段１０３はロック／メモリ対応情報１０８とロック／退避領域対応情報１０１を参照し、図３および図４で示すロック識別子に対応するメモリアドレス３０２およびメモリサイズ３０３、退避領域情報４０２、退避領域情報（復旧用）４０３を取得し、ディスク読み込み手段１０６を使用してホスト間共有ディスク装置１４０のメモリ退避領域１４２から、ホスト内メモリ１０４にデータを読み込む。この時、メモリ退避領域１４２中の図５に示す書き込みチェック領域Ａ５０１と書き込みチェック領域Ｂ５０３をチェックし、ホストダウン等によりデータが不正な場合は、メモリ退避領域（復旧用）１４３から、ホスト内メモリ１０４にデータを読み込み、メモリ退避領域（復旧用）１４３の情報をメモリ退避領域１４２に書き込んだ後でユーザプログラム１０２に制御を戻す。メモリ退避領域１４２のデータが正しい場合はメモリ退避領域１４２の内容をディスク書き込み手段１０７を使用してメモリ退避領域（復旧用）１４３に書き込む。

【００５０】ユーザプログラム１０２から更新ロックのロック解除要求時に、クラスタ用ロック解除手段１０５はディスク書き込み手段１０７を使用してホスト内メモリ１０４をホスト間共有ディスク装置１４０のメモリ退避領域１４２に書き込み、ホスト間ロック管理手段１３０と連携し、ロック解除を行う。

【００５１】次に、あるホストでロック中にホストダウンが発生した場合の動作について説明する。

【００５２】（２１）参照ロックの場合

参照ロックは、共有メモリを排他的に参照することが目的であるため、メモリに対する復旧処理は不要である。

【００５３】ロックに関しては、ホスト１１０でホストダウンが発生した場合を例に説明する。ホスト稼動監視手段１２０は、ホスト１１０のダウンを認識すると、ホスト間ロック管理手段１３０に通知する。ホスト間ロック管理手段１３０はホスト１１０からのロックをすべて解除する。ロック解除により動作可能となったホスト１００のユーザプログラムは、前述の処理を行う。

【００５４】（２２）更新ロックの場合

更新ロックは、共有メモリを排他的に更新することが目的であるが、メモリに対する復旧処理は不要である。

【００５５】ロックの解除については、参照ロック時と同じしくみで動作する。ロック解除により動作可能となったホスト１００のユーザプログラムは、前述の処理を行う。ホスト間共有ディスク装置１４０のメモリ退避領域１４２にホスト内メモリ１０４のデータを書き込み中に障害が発生した場合でも、ロック時にデータのチェックが行われるため前回のロック解除成功時の正しいデータを参照することができる。

【００５６】次に、具体的な例を用いて説明する。

【００５７】（３１）ユーザプログラム１０２とユーザプログラム１１２がともに参照ロックの場合

ホスト１００のユーザプログラム１０２は、ロック／対応情報１０８を参照し、共有メモリを参照するためのロック（参照ロック）を行う。この時、ホスト間共有ディスク装置１４０中のメモリ退避領域１４２からホスト内メモリ１０４にデータが読み込まれる。次に、ホスト１１０のユーザプログラム１１２がロック／対応情報１１

８を参照し、共有メモリを参照するためのロック（参照ロック）を行う。参照ロックであるためロックに成功し、ユーザプログラム１０２と同様にホスト内メモリ１０４にデータが読み込まれる。

【００５８】（３２）ユーザプログラム１０２が参照ロックでユーザプログラム１１２が更新ロックの場合

ホスト１００のユーザプログラム１０２は、ロック／対応情報１０８を参照し、共有メモリを参照するためのロック（参照ロック）を行う。この時、ホスト間共有ディスク装置１４０中のメモリ退避領域１４２からホスト内メモリ１０４にデータが読み込まれる。次に、ホスト１１０のユーザプログラム１１２がロック／対応情報１１８を参照し、共有メモリを更新するためのロック（更新ロック）を行うとユーザプログラム１０２で参照ロック中であるためロック待ち状態となる。ユーザプログラム１０２でロックが解除された後でロックが成功し、ホスト間共有ディスク装置１４０中のメモリ退避領域１４２からホスト内メモリ１０４にデータが読み込まれる。ユーザプログラム１１２でロック解除時に、ホスト内メモリ１１４は、ホスト間共有ディスク装置１４０中のメモリ退避領域１４２に書き込まれる。

【００５９】（３３）ユーザプログラム１０２とユーザプログラム１１２がともに更新ロックの場合

ホスト１００のユーザプログラム１０２は、ロック／対応情報１０８を参照し、共有メモリを更新するためのロック（更新ロック）を行う。この時、ホスト間共有ディスク装置１４０中のメモリ退避領域１４２からホスト内メモリ１０４にデータが読み込まれる。次に、ホスト１１０のユーザプログラム１１２がロック／対応情報１１８を参照し、共有メモリを更新するためのロック（更新ロック）を行うとユーザプログラム１０２で更新ロック中であるためロック待ち状態となる。ユーザプログラム１０２がロック解除時に、ホスト内メモリ１０４は、ホスト間共有ディスク装置１４０中のメモリ退避領域１４２に書き込まれ、ロックが解除された後でユーザプログラム１１２でロックが成功し、ホスト間共有ディスク装置１４０中のメモリ退避領域１４２からホスト内メモリ１０４にデータが読み込まれる。ユーザプログラム１１２でロック解除時に、ホスト内メモリ１１４は、ホスト間共有ディスク装置１４０中のメモリ退避領域１４２に書き込まれる。

【００６０】

【発明の効果】以上説明したように、本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法は、第１の効果は、１ホストで構築していたシステム上で動作している多数のユーザプログラムを改造なしで複数ホストで構成されるクラスタシステムへ移行できることにある。

【００６１】その理由は、もし、ロック／アンロック機能と連動したメモリ共有ができない場合、１ホスト内で

しかユーザプログラムを実行できないが、本発明を利用することにより，ユーザプログラムを改造なしで複数ホストで実行できるからである。また、本発明を使用せず，ユーザプログラムをクラスタシステムに合った処理ロジックに改造する方法もあるが、その場合，多大な修正工数を要す。

【図面の簡単な説明】

【図１】本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法の一実施の形態を示すブロック図である。

【図２】本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるクラスタシステムを構成する１ホストであるメモリ共有型のマルチプロセスシステムを示すブロック図である。

【図３】本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるロック識別子と、そのロック識別子に対応する共有メモリであるホスト内メモリのメモリアドレスと、メモリサイズとの対応情報の構造を示す図である。

【図４】本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるロック識別子と、そのロック識別子に対応する共有メモリの退避領域情報と、退避領域情報（復旧用）との対応情報の構造を示す図である。

【図５】本発明の複数ホストから構成されるクラスタシステムにおけるメモリの共有装置と方法におけるホスト間共有ディスク装置のメモリ退避領域の書き込みチェック領域と、メモリ退避領域と、書き込みチェック領域との対応情報の構造を示す図である。

【符号の説明】

１００，１１０　ホスト
１０１　ロック／退避領域対応情報
１０２，１１２　ユーザプログラム
１０３，１１３　クラスタ用ロック手段
１０４，１１４　ホスト内メモリ
１０５，１１５　クラスタ用ロック解除手段
１０６，１１６　ディスク読み込み手段
１０７，１１７　ディスク書き込み手段
１０８，１１８　ロック／メモリ対応情報
１２０　ホスト稼動監視手段
１３０　ホスト間ロック管理手段
１４０　ホスト間共有ディスク装置
１４１　ロック／退避領域一元情報
１４２　メモリ退避領域
１４３　メモリ退避領域（復旧用）
３０１，４０１　ロック識別子
３０２　メモリアドレス
３０３　メモリサイズ
４０２　退避領域情報
４０３　退避領域情報（復旧用）

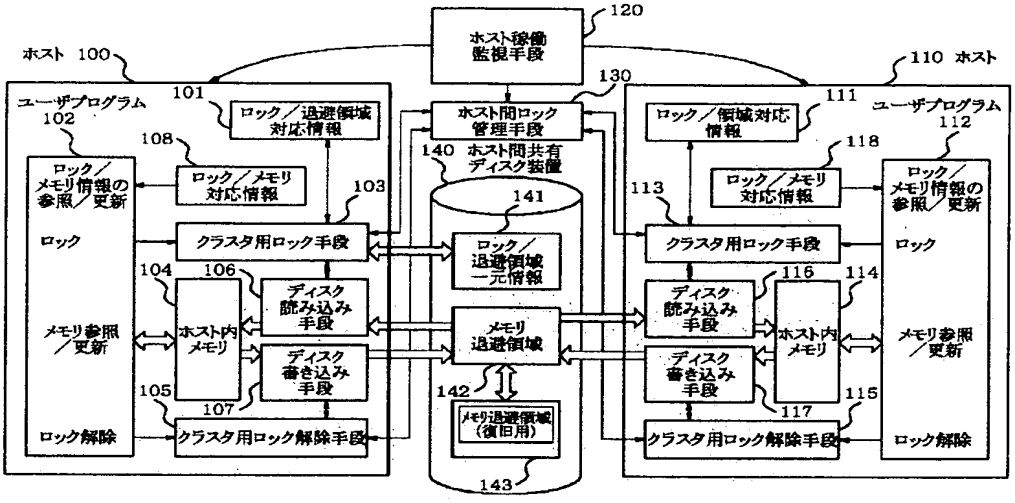
15

５０１　書き込みチェック領域Ａ
５０２　メモリ退避領域

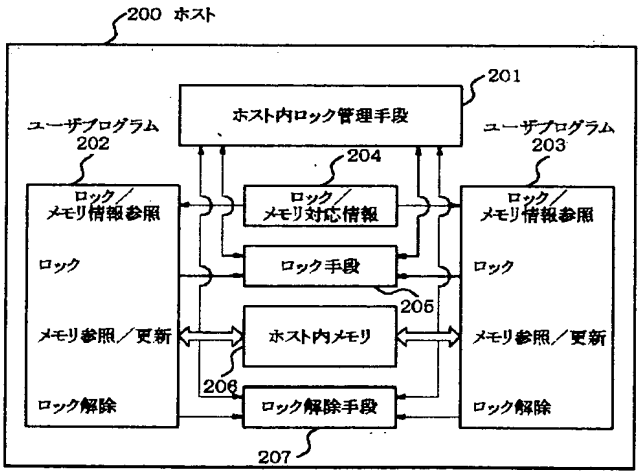
16

＊５０３　書き込みチェック領域Ｂ

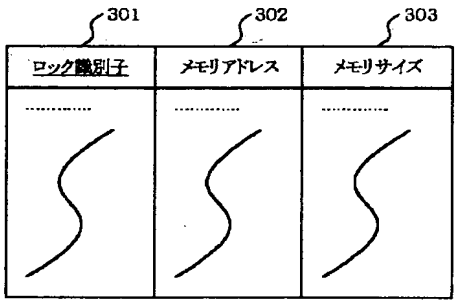
【図１】



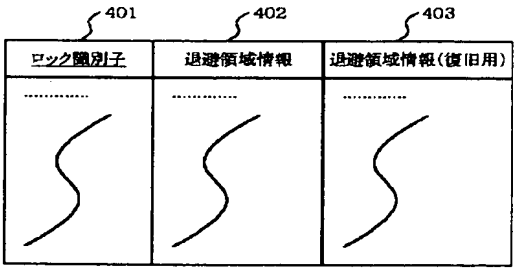
【図２】



【図３】



【図４】



【図５】